

# An Ergodic Theorem on Ergodic Transport

Joana Mohr and Rafael Rigão Souza

September 22, 2015

## Abstract

Here we present an ergodic theorem which adapts the setting of iterated function systems (IFS) for the ergodic transport theory. We will adapt a Theorem by J. Elton [2] to the classical thermodynamical formalism and also to ergodic transport. The main point here is to use an iterated function system to understand the main ergodic properties of transport plans (which solves a variational principle of ergodic transport as described in previous papers). To illustrate the technique, we first explore the thermodynamical formalism case, and only after that, we proceed to the ergodic transport case. We do that in the following way: after briefly recalling Elton's theorem, we discuss how it can be used to characterise Gibbs measures (also called equilibrium measures) for expanding maps, in the usual setting of thermodynamical formalism, where one is interested in finding the measure that maximizes the pressure of a given potential. Such characterisation will be done by constructing a stochastic process (defined by a IFS) whose empirical measure converges to the Gibbs measure, in the sense that the mean of any test function evaluated in the outcomes of this stochastic process converges to the integral of such test function with respect to the Gibbs measure. In this way we present a stochastic algorithm that compute integrals of functions. After this initial application, we turn our attention to ergodic transport: if we have two sets  $X$  and  $\Omega$ , a measure  $\mu$  on  $X$  and a dynamics  $T$  on  $\Omega$ , we can consider the set of probability measures on  $X \times \Omega$  whose projections on the second coordinate are  $T$ -invariant, while the projections in the first coordinate are given by  $\mu$ . Such measures are called transport plans. An optimal transport plan is a plan that maximizes the integral of a certain potential function. We also call Gibbs plan (or equilibrium transport plan) any transport plan that maximizes a pressure functional that is defined (as in the classical case) by a potential function plus an entropy term. Gibbs plan can converge, under certain conditions, to optimal plans (this is called the zero-temperature limit). We show how Elton's theorem can be used to characterise Gibbs plans in ergodic transport theory also by defining a stochastic process (defined by a IFS) whose empirical measures converges to the Gibbs plan, as in the classical thermodynamical formalism case. We provide examples and show explicitly calculations in the case where  $X$  has two elements and the cost function depends on the two first coordinates of  $\Omega$ .

## 1 Introduction

Ergodic transport can be motivated by an interesting problem in ergodic theory: given a dynamical system  $T : \Omega \rightarrow \Omega$ , where  $(\Omega, d)$  is a compact metric space,  $\mathcal{A}$  is the Borel sigma-algebra on  $\Omega$  and  $T$  is a continuous map, and given a fixed probability measure  $\mu$  on  $\mathcal{A}$  (which does not need to have any relation to the dynamics  $T$ ), one wants to obtain the  $T$ -invariant measure that minimizes the Wasserstein-2 distance to  $\mu$ .

If  $\nu$  is also a probability measure on  $\mathcal{A}$ , we denote by  $\Pi_{\mu, \nu}$  the set of probability measures on  $\Omega \times \Omega$  whose projections on each coordinates are, respectively,  $\mu$  and  $\nu$ . We have that the Wasserstein-2 distance between this measures is given by

$$W_2(\mu, \nu) = \sqrt{\inf_{\pi \in \Pi_{\mu, \nu}} \int d^2(x, y) d\pi(x, y)}.$$

It is known that this distance metrizes the weak-convergence topology on the space of probability measures on  $\mathcal{A}$  (see [9] for details).

Now, let  $\Pi_{\mu, T}$  be the set of probability measures on  $\Omega \times \Omega$  whose projections on the first coordinate are equal to  $\mu$  and the projections on the second coordinate are  $T$ -invariant. If one solves the constrained optimization problem

$$\inf_{\pi \in \Pi_{\mu, T}} \int d^2(x, y) d\pi(x, y),$$

one obtains (by projecting the solution on the second coordinate) the  $T$ -invariant measure that is closest to  $\mu$  (according to the Wasserstein metric). A measure  $\pi$  that solves the minimization problem above is called an optimal transport plan.

Ergodic transport was studied in a more general setting in [5]. Such papers consider probability measures on  $X \times \Omega$ , where  $X$  is a set that can be different from  $\Omega$ . If  $\mu$  is a probability measure on  $X$ , and  $T : \Omega \rightarrow \Omega$  is a dynamical system defined on  $\Omega$ , we denote by  $\Pi_{\mu,T}$  the set of probability measures on  $X \times \Omega$  that project (in the first coordinate) on  $\mu$ , while projecting, in the second coordinate, on any invariant measure for  $T : \Omega \rightarrow \Omega$ . See section 3 for the precise definition of  $\Pi_{\mu,T}$ . Note that different measures in  $\Pi_{\mu,T}$  can project in different invariant measures for  $T$ . In [5], one is interested in obtaining a measure that attains

$$\sup_{\pi \in \Pi_{\mu,T}} \int A(x,y) d\pi(x,y),$$

where  $A : X \times \Omega \rightarrow \mathbb{R}$  is called a potential function. If  $X$  is a singleton (has only one point), the maximization above is the same found in ergodic optimization problems, and for this reason we can see ergodic transport, in some sense, as a generalization of ergodic optimization. Note that, if we compare the problem above with the one concerning the measure that minimizes Wasserstein measure, we see that we are interested in maximize, rather than minimize, the integral of a function. This difference does not pose any conceptual differences in the theory: if one wants to minimize, he just have to change the sign of the potential.

After optimal plans were introduced, in [6] the entropy of plans was defined by means of what can be called the Jacobian of the measure. If we define a functional by adding the integral of the potential with respect to some plan, with the entropy of this plan, and try to maximize this functional, we have what is called a pressure problem (in analogy to classical thermodynamical formalism). One can think that we are considering Thermodynamic Formalism where the potential is random due to the choice of a fixed probability  $\mu$ . A variational principle was obtained in [6] (see section 3 for more details on the results of [6] that are needed here). The plan that satisfies the variational principle is called an equilibrium plan, and is related to the fixed point of a Ruelle-Perron-Frobenius-type operator (transfer operator). Moreover, as a relation between equilibrium plans and optimal plans, in [6] it is proved that optimal plans can be obtained as weak limits of equilibrium plans, if we multiply the cost by a constant  $\beta$ , and send  $\beta \rightarrow \infty$ . This is called the zero temperature limit (also in analogy to classical thermodynamical formalism - see [1]).

In section 3 we review the basic results in ergodic transport, and some details of the results stated above.

Section 4 brings the main objective of this paper, which is to characterize, via an Ergodic theorem, the equilibrium plan, that solves the pressure problem introduced in [6]. This Theorem plays the same role in Ergodic Transport as the Ergodic Theorem in classical Ergodic Theory.

Such characterisation will be done by constructing a stochastic process whose empirical measure converges to the Gibbs measure, in the sense that the mean of any test function evaluated in the outcomes of this stochastic process converges to the integral of the test function with respect to the equilibrium plan. We provide examples and also show explicitly calculations in the case where  $X$  has two elements and the cost function depends on the two first coordinates of  $\Omega$ .

The main tool in getting this characterization is a result due to Elton, concerning an ergodic theorem for Markov processes defined by iterated function systems (IFS), which is stated below in an adapted form that is appropriate for our purposes:

**Theorem 1 (Elton - 1987)** *Let  $Z$  be a compact metric space, and  $\tau_i : Z \rightarrow Z$ , for  $1 \leq i \leq d$  be a finite number of Lipschitz contractive maps. Let  $p_i : Z \rightarrow (0, 1]$  be Lipschitz continuous weight functions, and suppose  $\sum_{i=1}^d p_i(z) = 1$  for all  $z \in Z$ . Let  $\nu$  be a probability measure on  $Z$  that satisfies, for any Borel-measurable set  $B$ ,*

$$\nu(B) = \int_Z \sum_{\tau_i(z) \in B} p_i(z) d\nu(z). \quad (1)$$

*Then, for any  $z_0 \in Z$ , if we define by recurrence*

$$z_{k+1} = \tau_i(z_k) \quad \text{with probability } p_i(z_k), \quad (2)$$

we have that, for any continuous function  $f : Z \rightarrow \mathbb{R}$ , almost surely,

$$\frac{1}{N} \sum_{k=0}^{N-1} f(z_k) \rightarrow \int f d\nu. \quad (3)$$

We will soon explain exactly what almost sure convergence in (3) means. Before that, some considerations are necessary: a finite set of maps  $\tau_i$  defined on  $Z$ , together with the *transition probabilities*  $p_i : Z \rightarrow (0, 1]$  defines what is called an iterated function system (IFS). In this paper we will call the maps  $\tau_i$  as Elton maps.

If we define the transition Kernel

$$P(z, B) = \sum_{\tau_i(z) \in B} p_i(z),$$

(1) means that  $\nu$  is the invariant measure for the Markov process  $\{z_k\}_{k \in \mathbb{N}}$  defined on  $Z$  by (2).

Elton's theorem implies that the law of large numbers holds for the process  $f(z_k)$ .

Note that the initial  $z_0 \in Z$  can be any point in  $Z$ : the time average on (3) does not depend on  $z_0$ .

Let us explain the meaning of almost sure convergence in (3): we begin by denoting  $\Omega = \{1, \dots, d\}^{\mathbb{N}}$  the Bernoulli set of  $d$  symbols, with the  $\sigma$ -algebra generated by the cylinders. According to [2], for any  $z \in Z$  there exists (i) a probability  $P_z$  on  $\Omega = \{1, \dots, d\}^{\mathbb{N}}$ , which is given in cylinders by

$$P_z(i_1, i_2, \dots, i_k) = p_{i_1}(z) p_{i_2}(\tau_{i_1}(z)) p_{i_3}(\tau_{i_2}(\tau_{i_1}(z))) \dots p_{i_k}(\tau_{i_{k-1}}(\dots(\tau_{i_1}(z)) \dots)),$$

(ii) a set  $G_z \in \Omega$  such that  $P_z(G_z) = 1$ . Now, [2] proves that, if the *address sequence*  $(i_1, i_2, \dots)$  belongs to  $G_z$ , then the sequence defined recursively by

$$\begin{cases} z_0 = z, \\ z_{k+1} = \tau_{i_{k+1}}(z_k), \text{ for } k \geq 0, \end{cases}$$

is such that (3) holds for any continuous function  $f : Z \rightarrow \mathbb{R}$ .

In J. Elton's result the orbits go backward and not forward as in the classical Birkhoff theorem.

As a simple application of Elton's result, let  $Z = \{1, 2, \dots, d\}$  and  $\tau_i = i$ . Then  $z_k$  defined by (2) is the usual Markov Chain associated to the transition matrix  $P_{ij} = p_j(i)$ . If  $P_{ij} > 0$  for all  $i, j$ , then Elton's theorem implies, as a particular case, the Law of Large numbers for the Markov chain  $z_k$ .

Elton's result was proved in a slightly more general form, where the maps  $\tau_i$ 's only need to be contractive 'on the average', and  $Z$  need not to be compact (see details in [2]). The invariant measure  $\nu$  that satisfies (1) was proved to be unique by some of the authors cited in [2]. More recently a very simple proof of Theorem 1 is provided in [3], in the case the  $p_i$ 's are constant and  $Z$  is compact.

Before applying Elton's result to characterize Equilibrium plans in ergodic transport, we will consider, in section 2, an analogous problem in classical thermodynamical formalism: we know that the action of the equilibrium measure on test functions can be obtained by a limit procedure considering pre-images of points (via Ruelle operator - see section 2). However, this procedure is not efficient in computational terms, because the number of preimages grows exponentially (see section 2). The method we propose, using Elton's theorem, is much more efficient because, using a Markovian stochastic process defined via an IFS, we get a law of large numbers that gives the integral of test functions as the limit of the mean of such functions evaluated in the outcomes of the process (see (3)).

## 2 Ergodic theorem in classical thermodynamical formalism

In this section we remember the basic facts of classical thermodynamical formalism and then use Elton's ergodic theorem to characterize Gibbs measures for the shift on the Bernoulli set of symbols. The results of this section are of independent interest, and will not be used in the following sections.

Let  $\Omega = \{1, \dots, d\}^{\mathbb{N}}$  be the Bernoulli set on  $d$  symbols, where  $d \in \mathbb{N}$ ,  $d \geq 2$ . We know (Tychonoff's theorem) that  $\Omega$  is a compact (metric) set. We will consider the sigma-algebra generated by the cylinders, which is the Borel sigma-algebra. The dynamics  $T$  here is given by the shift map on  $\Omega$ .

The Ruelle-Perron-Frobenius (see [4, 7, 8]) operator (also known as transfer operator) associated to a Lipschitz *potential*  $A : \Omega \rightarrow \mathbb{R}$  is the operator that associates to any continuous function  $\varphi : \Omega \rightarrow \mathbb{R}$  the continuous function  $L_A(\varphi)$  given by

$$L_A(\varphi)(z) = \sum_{T(w)=z} e^{A(w)} \varphi(w).$$

A Lipschitz function  $A : \Omega \rightarrow \mathbb{R}$  is called a *normalized potential* if  $L_A(1) = 1$ . It is known (see [8]) that the RPF operator has a maximal eigenvalue  $\lambda_A > 0$  associated to an eigenfunction  $\varphi_A$ , which is Lipschitz, simple and positive (a simple eigenvalue means an eigenvalue that has an associated eigenspace with dimension 1). If  $A$  is non-normalized then it can be normalized by considering

$$\bar{A} = A + \log \varphi_A - \log \varphi_A \circ T - \log \lambda_A.$$

The dual RPF operator, denoted by  $L_A^*$ , acts on probability measures on  $\Omega$ , and is defined by

$$\int \psi dL_A^*(\mu) = \int L_A(\psi) d\mu.$$

If  $\bar{A}$  is the normalized potential associated to  $A$ , we know the dual operator  $L_{\bar{A}}^*$  preserves the simplex of probabilities and therefore has a unique fixed probability  $\mu_A$  (i.e. a probability  $\mu_A$  that satisfies  $L_{\bar{A}}^*(\mu_A) = \mu_A$ ), called the Gibbs state associated to  $A$ , which is invariant and ergodic for  $T$ , and satisfies

$$\int A d\mu_A + h(\mu_A) = \sup_{\mu \in \mathcal{M}_T(\Omega)} \left\{ \int A d\mu + h(\mu) \right\}.$$

( $\mathcal{M}_T(\Omega)$  here means the set of invariant probabilities for  $T$ .) The right side of the equation above is called the pressure of  $A$ . We know that  $\mu_A$  is the unique invariant measure to attain the maximum defining the pressure. The unique measure that attains such maximum is called the equilibrium measure for  $A$ , is ergodic, gives positive mass to open sets, and is given by the Gibbs measure  $\mu_A$ . The last result is also known as the *variational principle for pressure*.

Now an interesting question is: how can we characterize the Gibbs measure  $\mu_A$ ?

More precisely, is it possible to calculate  $\int u d\mu_A$  for any test function  $u : \Omega \rightarrow \mathbb{R}$ ?

A partial answer is given by the fact (see [8]) that, for any Lipschitz test function  $u : \Omega \rightarrow \mathbb{R}$ , we have

$$\frac{L_A^n(u)}{\lambda_A^n} \rightarrow \varphi_A \int u d\mu_A, \quad (4)$$

when  $n \rightarrow \infty$ , where

$$L_A^n(\varphi)(z) = \sum_{T^n(w)=z} e^{S_n(A)(w)} \varphi(w),$$

where  $S_n(A) = \sum_{k=0}^{n-1} A \circ T^k$  is the Birkhoff sum of order  $n$  of  $A$ . The convergence above is on the uniform convergence topology.

However, this limit procedure above is not of practical use because it involves the evaluation of  $u$  in the inverse images of order  $n$ , with  $n \rightarrow \infty$ , of some point of  $\Omega$ , and such calculation is not practical to be implemented: for the shift on  $d$  symbols, we have  $d^n$  inverse images of order  $n$ . To get things even worse, we have to evaluate the Birkhoff sum of the potential  $A$  of order  $n$  in each one of the  $d^n$  inverse images of a chosen point.

Therefore, it is necessary another way for calculating  $\int u d\mu_A$ . This can be accomplished by using Elton's theorem. In this way we get a much more efficient procedure, because what we will get is a Markovian stochastic process (that can be easily simulated in a Monte-Carlo process) and we will get a law of large numbers that will result in the integral of any test function, as shown in equation (3).

Suppose  $A : \Omega \rightarrow \mathbb{R}$  is normalized, i. e.  $\sum_{i=1}^d e^{A(iz)} = 1 \quad \forall z \in \Omega$ .

Let  $\tau_i(z) = iz$  be the Elton maps, and the transition probabilities be given by  $p_i(z) = e^{A(iz)}$ . The normalization hypothesis on  $A$  implies that  $\sum p_i(z) = 1$  for any  $z$ , and also the Lipschitz continuity of  $A$  implies that  $p_i$  is Lipschitz.

Let  $\mu$  be the probability measure that satisfies (1) in Elton's theorem. Let  $B$  be any measurable set of  $\Omega$ :

Then

$$\begin{aligned}\mu(B) &= \int \sum_{\tau_i(z) \in B} p_i(z) d\mu(z) = \int \sum_{iz \in B} e^{A(iz)} d\mu(z) = \\ &= \int \sum_i e^{A(iz)} I_B(iz) d\mu(z) = \int \mathcal{L}_A(I_B) d\mu = \mathcal{L}_A^* \mu(B),\end{aligned}$$

and therefore  $\mu$  is the fixed point of Ruelle dual operator.

As a result, Elton measure  $\mu$  coincides with the Gibbs measure  $\mu_A$ , and we have

**Theorem 2 (Birkhoff-Elton Theorem in classical thermodynamical formalism)** *If we choose any  $z_0 \in \Omega$ , and then, by recurrence, choose*

$$z_{k+1} = iz_k \text{ with probability } p_i(z_k) = e^{A(iz_k)},$$

*then, for any continuous test function  $f : \Omega \rightarrow \mathbb{R}$ ,*

$$\frac{1}{N} \sum_{k=0}^{N-1} f(z_k) \rightarrow \int f d\mu_A,$$

*with probability one.*

### 3 Ergodic Transport

Now we recall the main concepts of Ergodic Transport. See [6] for more details.

Let  $X$  be a finite set and  $\Omega = \{1, \dots, d\}^{\mathbb{N}}$  the Bernoulli space on  $d$  symbols.

We denote the RPF operator associated to a Lipschitz cost (potential)  $c : X \times \Omega \rightarrow \mathbb{R}$  as

$$L_c(v)(y) = \sum_{x \in X} \sum_{\sigma(w)=y} e^{c(x,w)} v(w) = \sum_{\sigma(w)=y} \left( \sum_{x \in X} e^{c(x,w)} \right) v(w),$$

for any  $v \in C(\Omega)$ . We remark that this is the classical RPF operator associated to the potential  $b_c(y) = \log(\sum_{x \in X} e^{c(x,y)})$ . We know that  $L_c$  has a maximal eigenvalue  $\lambda_c$ , which is simple and positive, and there is a positive eigenfunction  $h_c$  associated to  $\lambda_c$ , see [8].

The RPF extended to the continuous functions  $X \times \Omega$  is defined as

$$\hat{L}_c(u)(y) = \sum_{x \in X} \sum_{\sigma(w)=y} e^{c(x,w)} u(x, w),$$

for any  $u : X \times \Omega \rightarrow \mathbb{R}$ .

Note that  $\hat{L}_c$  sends  $u : X \times \Omega \rightarrow \mathbb{R}$  to the function denoted by  $\hat{L}_c(u) : \Omega \rightarrow \mathbb{R}$ .

**Definition 1** *We say that a Lipschitz cost (potential)  $c : X \times \Omega \rightarrow \mathbb{R}$  is normalized if for any  $y \in \Omega$ , we have*

$$\sum_{x \in X} \sum_{\sigma(w)=y} e^{c(x,w)} = 1.$$

If  $c$  is a Lipschitz cost that is not normalized, we associate to  $c$  the normalized cost

$$\bar{c}(x, y) = c(x, y) + \log(h_c(y)) - \log(h_c \circ \sigma(y)) - \log(\lambda_c), \quad (5)$$

where  $h_c$  is the positive eigenfunction associated to the maximal eigenvalue  $\lambda_c$  of  $L_c$ .

If  $c$  is normalized, we denote  $\hat{L}_c^*$  the operator on  $P(X \times \Omega)$  defined by

$$\hat{L}_c^*(\pi)(u(x, y)) = \int_{X \times \Omega} \left( \sum_{\alpha \in X} \sum_{\sigma(w)=y} e^{c(\alpha, w)} u(\alpha, w) \right) d\pi(x, y). \quad (6)$$

The normalization property implies that  $\hat{L}_c^*(\pi)(1) = 1$ , i.e.  $\hat{L}_c^*$  preserves the convex and compact set  $P(X \times \Omega)$ , and Tychonoff-Schauder theorem implies

**Proposition 1** *Given a normalized cost  $c$  there exists a unique fixed point for the operator  $\hat{L}_c^*$ . It will be denoted by  $\pi_c$ .*

Therefore, we have

$$\pi_c(u(x, y)) = \int_{X \times \Omega} \left( \sum_{\alpha \in X} \sum_{\sigma(w)=y} e^{c(\alpha, w)} u(\alpha, w) \right) d\pi_c(x, y). \quad (7)$$

**Definition 2** *The fixed point,  $\pi_c$ , for  $\hat{L}_c^*$  is called the **Gibbs plan** for the normalized cost (potential)  $c$ .*

We denote by  $\Pi(\cdot, \sigma)$  the set of plans such that its  $y$ -marginal is  $\sigma$ -invariant, i.e.,

$$\int_{X \times \Omega} g(y) d\pi(x, y) = \int_{X \times \Omega} g(\sigma(y)) d\pi(x, y) \quad \text{for any } g \in C(\Omega). \quad (8)$$

By Theorem 4 of [6], we have that  $\pi_c \in \Pi(\cdot, \sigma)$  and the  $y$ -marginal of  $\pi_c$  is the Gibbs measure  $\nu_c$  to the classical RPF operator with the potential  $b_c(y) = \log(\sum_{x \in X} e^{c(x, y)})$ .

We denote by  $[x, y_1 \dots y_n] = \{(\alpha, w) \in X \times \Omega : \alpha = x, w_1 = y_1, \dots, w_n = y_n\}$  and  $[y_2 \dots y_n] = \{w \in \Omega : w_1 = y_2, \dots, w_{n-1} = y_n\}$ . Consider a fixed plan  $\pi \in \Pi(\cdot, \sigma)$  with  $y$ -marginal  $\nu$  and define

$$J_\pi^n(x, y) = \frac{\pi([x, y_1 \dots y_n])}{\nu([y_2 \dots y_n])}$$

if  $y = (y_2, y_3, \dots) \in \text{supp}(\nu)$ . From the Increasing Martingale Theorem the functions  $J_\pi^n$  converge to a function  $J_\pi(x, y)$  in  $L^1(X \times \Omega, \mathcal{A}(X \times \Omega), \pi)$  and for  $\pi$  a.e.  $(x, y)$ . For each plan  $\pi$  this function  $J_\pi$  can be also obtained via the Radon-Nikodym Theorem. We have,  $J_\pi > 0$  a.e.  $(\pi)$  and  $\sum_{x \in X} \sum_{a \in \{1, \dots, d\}} J_\pi(x, ay) = 1$ .

We define the entropy of a plan  $\pi$  as

$$H(\pi) = - \int \log(J_\pi) d\pi.$$

**Definition 3** *The pressure of a Lipschitz continuous cost (potential)  $c$  is defined by*

$$P(c) = \sup_{\pi \in \Pi(\cdot, \sigma)} \left( \int_{X \times \Omega} c d\pi + H(\pi) \right).$$

A plan  $\pi \in \Pi(\cdot, \sigma)$  which realizes the supremum is called **an equilibrium plan** for  $c$ .

**Theorem 3 (Variational Principle over  $\Pi(\cdot, \sigma)$ )** *Let us fix a Lipschitz cost  $c$ . Then,  $P(c) = \log(\lambda_c)$ , where  $\lambda_c$  is the main eigenvalue of  $L_c$ . The equilibrium plan for  $c$  is unique and given by the Gibbs plan for the normalized cost  $\bar{c} := c + \log(h_c) - \log(h_c \circ \sigma) - \log(\lambda_c)$ , where  $h_c$  is the positive eigenfunction associated to  $\lambda_c$ .*

Now let us fix  $\mu$  a probability on  $X$ . We define the  $\mu$ -pressure of  $c$  by

$$P_\mu(c) = \sup_{\pi \in \Pi(\mu, \sigma)} \int_{X \times \Omega} c d\pi + H(\pi), \quad (9)$$

where  $\Pi(\mu, \sigma)$  is the set of all plans satisfying

$$\begin{cases} \int_{X \times \Omega} f(x) d\pi(x, y) = \int_X f(x) d\mu(x) & \text{for any } f \in C(X), \\ \int_{X \times \Omega} g(y) d\pi(x, y) = \int_{X \times \Omega} g(\sigma(y)) d\pi(x, y) & \text{for any } g \in C(\Omega), \end{cases} \quad (10)$$

which means, the set of probabilities  $\pi$  such that the  $x$ -marginal of  $\pi$  is the fixed probability  $\mu \in P(X)$  and the  $y$ -marginal of  $\pi$  is  $\sigma$ -invariant.

Note that  $P_\mu(c) \leq P(c)$ .

By compactness, there exists a plan  $\tilde{\pi}_c \in \Pi(\mu, \sigma)$  which attains the supremum at (9).

We have the following duality result:

**Theorem 4** *Given a Lipschitz cost  $c$ , and let  $\mu$  be a probability on  $X$ , we have*

$$\inf_{\varphi: P(c-\varphi)=0} \int_X \varphi d\mu = \sup_{\pi \in \Pi(\mu, \sigma)} \int_{X \times \Omega} c(x, y) d\pi + H(\pi). \quad (11)$$

*The minimization is performed on the set of functions  $\varphi : X \rightarrow \mathbb{R}$  such that  $P(c - \varphi) = 0$ . The supremum at (11) is attained in at least one plan, while the infimum is attained in exactly one function  $\tilde{\varphi}$ .*

Moreover, if  $\tilde{\varphi} : X \rightarrow \mathbb{R}$  is the unique minimizer of (11), then  $P(c - \tilde{\varphi}) = 0$  and also

$$\int_X \tilde{\varphi} d\mu = \sup_{\pi \in \Pi(\mu, \sigma)} \int_{X \times \Omega} c(x, y) d\pi + H(\pi)$$

which implies

$$\sup_{\pi \in \Pi(\mu, \sigma)} \int_{X \times \Omega} (c(x, y) - \tilde{\varphi}) d\pi + H(\pi) = 0,$$

which means  $P_\mu(c - \tilde{\varphi}) = 0$ .

Therefore  $P(c - \tilde{\varphi}) = P_\mu(c - \tilde{\varphi}) = 0$ , and the maximizer of (11) is the equilibrium plan for  $c - \tilde{\varphi}$ , i.e. satisfies the non-constrained variational principle. We just proved the

**Corollary 1** *Let  $\tilde{\varphi} : X \rightarrow \mathbb{R}$  be the unique minimizer for the Fenchel-Rockafellar duality (11). The equilibrium plan  $\pi_{c-\tilde{\varphi}}$ , for  $c - \tilde{\varphi}$ , belongs to  $\Pi(\mu, \sigma)$  and is the unique maximizer of (11).*

As a final remark in this section, let us observe that we show, in [6], among other things, how equilibrium plans can be used to obtain the optimal transport plan, by a limit procedure where the cost  $c$  is multiplied by a constant  $\beta$  (the inverse of the temperature) and the parameter  $\beta \rightarrow \infty$ .

## 4 An ergodic theorem in ergodic transport

Now we discuss the main goal of this paper: the characterization, via Elton theorem, of the plan  $\pi_{c, \mu}$  that satisfies the supremum

$$\sup_{\pi \in \Pi(\mu, \sigma)} \int_{X \times \Omega} c(x, y) d\pi + H(\pi).$$

where  $X = \{1, 2\}$ ,  $\mu$  is a probability measure on  $X$ ,  $c : X \times \Omega \rightarrow \mathbb{R}$  depends on  $x \in X$  and only on the two first coordinates of  $y \in \Omega$ , and  $\Pi(\mu, \sigma)$  is given by (10). Note that we are dealing with the **constrained** optimization problem, i.e. we search an optimising measure among those measures that project on a fixed probability measure  $\mu$  on  $X$ . Any such measure is given by  $\mu = (p, 1 - p)$ , where  $0 \leq p \leq 1$ .

By the corollary 1, the plan  $\pi_{c, \mu}$  is the equilibrium plan  $\pi_{c-\tilde{\varphi}}$ , where  $\tilde{\varphi}$  is the minimizer of (11). Note that this minimization is performed on the set of functions  $\varphi : X \rightarrow \mathbb{R}$  such that  $P(c - \varphi) = 0$ .

We will do such characterisation in three steps:

**Step 1** We need to find the minimizer  $\tilde{\varphi} = (\varphi_1, \varphi_2) \in \mathbb{R}^2$  that solves the minimization problem (11), i.e. that solves

$$\inf_{\varphi: P(c-\varphi)=0} \int_X \varphi d\mu.$$

Such minimizer exists and is unique (theorem 4).

**Step 2** If  $A = c - \tilde{\varphi}$ , we need to find the maximal eigenvalue  $\lambda_A$  for the RPF operator  $L_A$ , as well as the associated eigenfunction  $h_A$ . Once we have that, we define the normalized potential  $\bar{A} = A + \log(h_A) - \log(h_A \circ \sigma) - \log(\lambda_A)$ , which can be reduced to

$$\bar{A} = A + \log(h_A) - \log(h_A \circ \sigma),$$

as  $\log(\lambda_A) = P(A) = P(c - \bar{\varphi}) = 0$ .

**Step 3** We use Elton Theorem to characterize the constrained equilibrium measure  $\pi_{c,\mu}$  showing how one can get the integral of any test function by means of a stochastic simulation (a Monte Carlo method). More precisely, for any continuous function  $f : X \times \Omega \rightarrow \mathbb{R}$ , show that, almost certainly,

$$\frac{1}{N} \sum_{k=0}^{N-1} f(x^k, y^k) \rightarrow \int f d\pi_{c,\mu},$$

for a well-choosen sequence  $(x^k, y^k)$ , obtained by a probabilistic algorithm.

### Analysis of Step 3:

Now we suppose Steps 1 and 2 were successfully performed (the analysis of these two first steps will be done after theorem 5). We proceed to step 3:

By Corollary 1 we know that  $\pi_{c,\mu}$  is the equilibrium measure  $\pi_A$  or  $\pi_{c-\bar{\varphi}}$ , and by theorem 3 we have that  $\pi_A$  or  $\pi_{c-\bar{\varphi}}$  is the Gibbs plan for  $\bar{A}$ , i.e.,

$$\hat{L}_A^*(\pi_A) = \pi_A.$$

Now we want to show that  $\pi_A$  (i.e.  $\pi_{c,\mu}$ ) satisfies the hypothesis of theorem 1.

Let us take, following the notation used in Elton's Theorem 1,  $Z = X \times \Omega$ , then we define for all  $(\alpha, i) \in X \times \{1, \dots, d\}$  and for all  $(x, y) \in Z$ , the weight function

$$p_{\alpha,i}(x, y) = e^{\bar{A}(\alpha, iy)}.$$

Note that  $p_{\alpha,i}$  depends only on  $y$ .

As  $\bar{A}$  is normalized we have

$$1 = \sum_{\alpha \in X} \sum_{1 \leq i \leq d} e^{\bar{A}(\alpha, iy)} = \sum_{\alpha \in X, 1 \leq i \leq d} p_{\alpha,i}(x, y), \quad \forall (x, y) \in Z.$$

For each  $\alpha \in X, 1 \leq i \leq d$ , we define the Elton map

$$\tau_{\alpha,i}(x, y) = (\alpha, iy).$$

Finally, if  $B$  is a borelian set of  $Z = X \times \Omega$ , and  $\chi_B$  is the characteristic function of  $B$ . We have, by equation (7) that

$$\begin{aligned} \pi_A(B) &= \int \hat{L}_A(\chi_B) d\pi_A = \int \sum_{\alpha \in X, 1 \leq i \leq d} p_{\alpha,i}(x, y) \chi_B(\alpha, iy) d\pi_A(x, y) \\ &= \int \sum_{\tau_{\alpha,i}(x, y) \in B} p_{\alpha,i}(x, y) d\pi_A(x, y), \end{aligned}$$

which means that  $\pi_A$  satisfies (1) and the Elton Theorem is true for the stochastic process generated by  $p_{\alpha,i}$  and  $\tau_{\alpha,i}$ . Therefore, we have proved the following theorem:

**Theorem 5** [Birkhoff-Elton Theorem in Ergodic Transport] Fix any  $(x^0, y^0) \in X \times \Omega$ . Define by recurrence

$$(x^{k+1}, y^{k+1}) = (\alpha, iy^k) \text{ with probability } e^{\bar{A}(\alpha, iy^k)},$$

where  $\bar{A}$  is the normalized cost associated to  $A = c - \bar{\varphi}$ , and  $\bar{\varphi}$  is the only minimizer of the Fenchel-Rockafellar duality equation (11). Then, for any continuous function  $f : X \times \Omega \rightarrow \mathbb{R}$ , we have, almost certainly,

$$\frac{1}{N} \sum_{k=0}^{N-1} f(x^k, y^k) \rightarrow \int f d\pi_A,$$

where  $\pi_A = \pi_{c,\mu}$  is the constrained maximizer of the duality equation (11).



### Analysis of Steps 1 and 2:

We will consider here the case  $d = 2$ , i.e.,  $\Omega = \{1, 2\}^{\mathbb{N}}$ . As  $c$  depends on  $x$  and only on the two first coordinates of  $y$ , that is,  $c(x, y) = c(x, y_1, y_2)$ , it is represented by the following matrices:

$$C^1 = \begin{pmatrix} e^{c_{11}^1} & e^{c_{12}^1} \\ e^{c_{21}^1} & e^{c_{22}^1} \end{pmatrix}, \quad C^2 = \begin{pmatrix} e^{c_{11}^2} & e^{c_{12}^2} \\ e^{c_{21}^2} & e^{c_{22}^2} \end{pmatrix}, \quad (12)$$

where  $C_{ij}^x = e^{c_{ij}^x} = e^{c(x, i, j)}$ ,  $x, i, j = 1, 2$ .

Note that the operator  $L_c : C(\Omega) \rightarrow C(\Omega)$ , for any  $v : \Omega \rightarrow \mathbb{R}$ , can be rewritten as

$$L_c(v)(y) = \sum_{i \in \{1, 2\}} \left( \sum_{x \in X} e^{c(x, iy)} \right) v(iy) = \sum_{i \in \{1, 2\}} e^{b_c(iy)} v(iy),$$

where  $b_c(y) = \log \left( \sum_{x \in X} e^{c(x, y)} \right)$ . Then  $b_c : \Omega \rightarrow \mathbb{R}$  depends on the two first coordinates of  $y$  and is represented by the matrix

$$B = \begin{pmatrix} e^{c_{11}^1} + e^{c_{11}^2} & e^{c_{12}^1} + e^{c_{12}^2} \\ e^{c_{21}^1} + e^{c_{21}^2} & e^{c_{22}^1} + e^{c_{22}^2} \end{pmatrix}. \quad (13)$$

It is a well known fact, see [8], that the positive eigenfunction, that we denote by  $v_c$ , associated to the main eigenvalue,  $\lambda_c$ , of  $L_c$  depends only on the first coordinate of  $y$ , and this means that  $v_c$  is in fact a vector of  $\mathbb{R}^2$ .

More precisely,  $v_c$  is the left eigenvector of  $B$  associated to the eigenvalue  $\lambda_c$ , because

$$L_c(v_c)(y) = \lambda_c v_c(y) \iff v_c B = \lambda_c v_c. \quad (14)$$

Let also

$$C^{12} = \begin{pmatrix} e^{c_{11}^1} & e^{c_{12}^1} \\ e^{c_{21}^2} & e^{c_{22}^2} \end{pmatrix}, \quad C^{21} = \begin{pmatrix} e^{c_{11}^2} & e^{c_{12}^2} \\ e^{c_{21}^1} & e^{c_{22}^1} \end{pmatrix}.$$

The next Theorem shows how to perform the Step 1.

**Theorem 6** *Let  $\mu = (p, 1 - p)$ . Suppose  $c : X \times \Omega \rightarrow \mathbb{R}$  is a Lipschitz cost that depends on  $x$  and just on the two first coordinates of  $y$ .*

(a) *Then the unique minimizer  $\tilde{\varphi}$  of the left-hand side of the equation*

$$\inf_{\varphi: P(c-\varphi)=0} \int_X \varphi(x) d\mu = \sup_{\pi \in \Pi(\mu, \sigma)} \int_{X \times \Omega} c d\pi + H(\pi) \quad (15)$$

is given by

$$\begin{cases} \tilde{\varphi}_1 = -\log(z_1), \\ \tilde{\varphi}_2 = -\log(z_2), \end{cases}$$

where  $(z_1, z_2)$  belongs to the set of positive solutions of the following system of equations

$$\begin{cases} a z_1^2 + b z_2^2 + c z_1 z_2 + d z_1 + e z_2 + 1 = 0, \\ 2bpz_2^2 - 2a(1-p)z_1^2 + c(2p-1)z_1 z_2 + epz_2 - d(1-p)z_1 = 0, \end{cases} \quad (16)$$

where  $a = \det C^1$ ,  $b = \det C^2$ ,  $c = \det C^{12} + \det C^{21}$ ,  $d = -\text{tr} C^1$ , and  $e = -\text{tr} C^2$ .

This means that  $(z_1, z_2)$  belongs to the set of intersection points of the two conics given by the equations above. Such intersections points are four, at most, and  $(z_1, z_2)$  is the one that minimizes  $-p \log(z_1) - (1-p) \log(z_2)$ .

(b) Suppose also that  $C^1$  and  $C^2$  are stochastic matrices. Then the unique minimizer  $\tilde{\varphi}$  of the left-hand side of the equation (15) is given by

$$\begin{cases} \tilde{\varphi}_1 = -\log(p), \\ \tilde{\varphi}_2 = -\log(1-p). \end{cases}$$

Also we have

$$\sup_{\pi \in \Pi(\mu, \sigma)} \int_{X \times \Omega} c d\pi + H(\pi) = -p \log(p) - (1-p) \log(1-p).$$

**Proof:** First we need to understand what means the restriction,  $P(c - \varphi) = 0$ , in left-hand side of equation (15). By Theorem 3  $P(c - \varphi) = \log(\lambda_{c-\varphi})$ , where  $\lambda_{c-\varphi}$  is the main eigenvalue of  $L_{c-\varphi}$ . Therefore  $P(c - \varphi) = 0$  means  $\lambda_{c-\varphi} = 1$ . As  $\varphi : X \rightarrow \mathbb{R}$  and  $X = \{1, 2\}$ ,  $\varphi$  can be seen as two-dimensional vector  $\varphi = (\varphi_1, \varphi_2)$ . Then we have

$$L_{c-\varphi}\psi(y) = \sum_i \left[ e^{c(1,ij)-\varphi(1)} + e^{c(2,ij)-\varphi(2)} \right] \psi(iy),$$

where  $y = (j, y_2, y_3, \dots)$ . As in (13), the operator  $L_{c-\varphi}$  is represented by the  $2 \times 2$  matrix

$$B = \begin{pmatrix} e^{c_{11}^1} e^{-\varphi_1} + e^{c_{11}^2} e^{-\varphi_2} & e^{c_{12}^1} e^{-\varphi_1} + e^{c_{12}^2} e^{-\varphi_2} \\ e^{c_{21}^1} e^{-\varphi_1} + e^{c_{21}^2} e^{-\varphi_2} & e^{c_{22}^1} e^{-\varphi_1} + e^{c_{22}^2} e^{-\varphi_2} \end{pmatrix} \quad (17)$$

finally using (14),  $P(c - \varphi) = 0$  means  $\lambda_{c-\varphi} = 1$  is the dominant eigenvalue of  $B$ .

If we apply the change of coordinates  $z_1 = e^{-\varphi_1}$ ,  $z_2 = e^{-\varphi_2}$ . Then  $P(c - \varphi) = 0$  if and only if  $(z_1, z_2)$  satisfy

(i)  $z_1 > 0, z_2 > 0$ ,

(ii)

$$\det \begin{pmatrix} e^{c_{11}^1} z_1 + e^{c_{11}^2} z_2 - 1 & e^{c_{12}^1} z_1 + e^{c_{12}^2} z_2 \\ e^{c_{21}^1} z_1 + e^{c_{21}^2} z_2 & e^{c_{22}^1} z_1 + e^{c_{22}^2} z_2 - 1 \end{pmatrix} = 0,$$

(iii) the other eigenvalue of the matrix  $B$  given in (17) is less than 1.

Note that we want to find the unique minimizer of

$$\inf_{\varphi: P(c-\varphi)=0} \int_X \varphi(x) d\mu,$$

but instead of that we will minimize  $\int_X \varphi(x) d\mu$ , subject to the restriction that  $(z_1, z_2)$  satisfies condition (ii)<sup>1</sup>. After we get all the solutions of this problem we will test which of them also satisfies conditions (i) and (iii).

Condition (ii) is equivalent to

$$\begin{aligned} & (e^{c_{11}^1} z_1 + e^{c_{11}^2} z_2 - 1)(e^{c_{22}^1} z_1 + e^{c_{22}^2} z_2 - 1) - (e^{c_{12}^1} z_1 + e^{c_{12}^2} z_2)(e^{c_{21}^1} z_1 + e^{c_{21}^2} z_2) = \\ & = z_1^2(e^{c_{11}^1} e^{c_{22}^1} - e^{c_{12}^1} e^{c_{21}^1}) + z_2^2(e^{c_{11}^2} e^{c_{22}^2} - e^{c_{12}^2} e^{c_{21}^2}) + z_1(-e^{c_{11}^1} - e^{c_{22}^1}) + z_2(-e^{c_{11}^2} - e^{c_{22}^2}) + \\ & \quad + z_1 z_2(e^{c_{11}^1} e^{c_{22}^2} + e^{c_{11}^2} e^{c_{22}^1} - e^{c_{12}^1} e^{c_{21}^2} - e^{c_{12}^2} e^{c_{21}^1}) + 1 = \\ & := a z_1^2 + b z_2^2 + c z_1 z_2 + d z_1 + e z_2 + 1 = 0. \end{aligned}$$

If we denote by

$$g(z_1, z_2) = a z_1^2 + b z_2^2 + c z_1 z_2 + d z_1 + e z_2 + 1$$

then condition (ii) describes an algebraic curve (a conic) on  $\mathbb{R}^2$ , given by

$$g(z_1, z_2) = 0.$$

Therefore we need to minimize

$$f(z_1, z_2) := \int_X \varphi(x) d\mu = \varphi_1 p + \varphi_2 (1 - p) = -p \log z_1 - (1 - p) \log z_2,$$

subject to the restriction  $g(z_1, z_2) = 0$ . To do that we will use the Lagrange multiplier theorem: if  $(z_1, z_2)$  is a solution to the constrained minimization problem above, there exists a  $\lambda \in \mathbb{R}$  such that

$$\nabla f(z_1, z_2) = \lambda \nabla g(z_1, z_2).$$

Using that

$$\begin{cases} \nabla g(z_1, z_2) = (2a z_1 + c z_2 + d, 2b z_2 + c z_1 + e), \\ \nabla f(z_1, z_2) = \left(-\frac{p}{z_1}, -\frac{1-p}{z_2}\right), \end{cases}$$

---

<sup>1</sup>see remark after Theorem 6

we need to solve the following equations

$$-\frac{p}{z_1} = \lambda(2az_1 + cz_2 + d) \quad \text{and} \quad -\frac{1-p}{z_2} = \lambda(2bz_2 + cz_1 + e),$$

or

$$\frac{p}{z_1(2az_1 + cz_2 + d)} = \frac{1-p}{z_2(2bz_2 + cz_1 + e)},$$

or

$$p(2bz_2^2 + cz_1z_2 + ez_2) = (1-p)(2az_1^2 + cz_2z_1 + dz_1),$$

which are equivalent to

$$2bpz_2^2 - 2a(1-p)z_1^2 + c(2p-1)z_1z_2 + epz_2 - d(1-p)z_1 = 0 \quad (18)$$

which, together with equation  $g = 0$  give the system of equations (16).

In order to conclude the proof of item (a) of Theorem 6, we remark that the unique minimizer of  $\inf_{\varphi: P(c-\varphi)=0} \int_X \varphi(x) d\mu$  belongs to the intersection of the two conics in (16), which are, at most, four points. For each of this solutions we test if it satisfies conditions (i), (that is, the solution have to be in the positive quadrant) and (iii). If there are more than one solution that satisfies conditions (i)-(iii), then we test which one is the minimizer.

Now, in order to prove item (b) of Theorem 6, let us suppose that  $C^1$  and  $C^2$  are stochastic matrices, i.e.

$$C^1 = \begin{pmatrix} e^{c_{11}^1} & e^{c_{12}^1} \\ 1 - e^{c_{11}^1} & 1 - e^{c_{12}^1} \end{pmatrix}, \quad C^2 = \begin{pmatrix} e^{c_{11}^2} & e^{c_{12}^2} \\ 1 - e^{c_{11}^2} & 1 - e^{c_{12}^2} \end{pmatrix} \quad (19)$$

We have  $a = e^{c_{11}^1} - e^{c_{12}^1}$ ,  $b = e^{c_{11}^2} - e^{c_{12}^2}$ ,  $c = a + b$ ,  $d = -(a + 1)$ ,  $e = -(b + 1)$ , and equation  $g = 0$  becomes

$$az_1^2 + bz_2^2 + (a+b)z_1z_2 - (a+1)z_1 - (b+1)z_2 + 1 = 0 \quad (20)$$

while equation (18) becomes

$$2bpz_2^2 - 2a(1-p)z_1^2 + (a+b)(2p-1)z_1z_2 - (b+1)pz_2 + (a+1)(1-p)z_1 = 0 \quad (21)$$

First we solve equation (20) in terms of  $z_2$ . We get, if  $a \neq 0$ , two solutions  $\tilde{z}_1 = \frac{1-bz_2}{a}$  and  $\bar{z}_1 = 1 - z_2$ , and if  $a = 0$  we get only the solution  $\bar{z}_1 = 1 - z_2$ .

Now, we solve (21):

Case 1: if  $\tilde{z}_1 = \frac{1-bz_2}{a}$ , (21) becomes

$$2bpz_2^2 - 2a(1-p)\left(\frac{1-bz_2}{a}\right)^2 + (a+b)(2p-1)\frac{1-bz_2}{a}z_2 - (b+1)pz_2 + (a+1)(1-p)\frac{1-bz_2}{a} = 0.$$

This equation can be rewritten as

$$\left(b - \frac{b^2}{a}\right)z_2^2 + \left(\frac{2b-bp}{a} + p - 1 - b\right)z_2 + \frac{p-1}{a} + (1-p) = 0$$

or

$$(ba - b^2)z_2^2 + (2b - bp + pa - a - ba)z_2 + p - 1 + a - pa = 0$$

finally, this is equivalent to

$$(a(z_2 - 1) - bz_2 + 1)(bz_2 + p - 1) = 0,$$

which have two solutions  $\tilde{z}_2 = \frac{a-1}{a-b}$  and  $\bar{z}_2 = \frac{1-p}{b}$ .

Then, we get two possible pair of solutions  $(z_1, z_2)$  for the system given by (20) and (21): if  $\tilde{z}_2 = \frac{a-1}{a-b}$ , then  $\tilde{z}_1 = \frac{1-bz_2}{a} = \frac{1-b}{a-b}$  and if  $\bar{z}_2 = \frac{1-p}{b}$ , then  $\bar{z}_1 = \frac{1-bz_2}{a} = \frac{p}{a}$ .

Case 2:  $\bar{z}_1 = 1 - z_2$ , then equation (21) becomes

$$2bpz_2^2 - 2a(1-p)(1-2z_2+z_2^2) + (a+b)(2p-1)(z_2-z_2^2) - (b+1)pz_2 + (a+1)(1-p)(1-z_2) = 0$$

which can be rewritten as

$$(-a+b)z_2^2 + (2a-ap+bp-b-1)z_2 + (ap-a-p+1) = 0.$$

Finally, this is equivalent to

$$(p+z_2-1)(a(z_2-1)-bz_2+1) = 0$$

which have two solutions  $\tilde{z}_2 = \frac{a-1}{a-b}$  and  $\bar{z}_2 = 1-p$ .

Then, we get another two possible pair of solutions  $(z_1, z_2)$ :

If  $\tilde{z}_2 = \frac{a-1}{a-b}$ , then  $\tilde{z}_1 = 1 - z_2 = \frac{1-b}{a-b}$  and if  $\bar{z}_2 = 1-p$ , then  $\bar{z}_1 = 1 - z_2 = p$ .

Collecting all the solutions of cases 1 and 2, we have 3 different solutions  $(z_1, z_2)$  for the system given by (20) and (21)

$$\left(\frac{1-b}{a-b}, \frac{a-1}{a-b}\right), \quad \left(\frac{p}{a}, \frac{1-p}{b}\right), \quad (p, 1-p).$$

Now we test conditions (i) and (iii) for each one of the three possible solutions above:

**Claim:**  $(\frac{1-b}{a-b}, \frac{a-1}{a-b})$  do not satisfy item (i), and  $(\frac{p}{a}, \frac{1-p}{b})$  do not satisfy item (i) or item (iii).

In fact, using that  $C^1$  and  $C^2$  are given by (19), we have that  $0 \leq e^{c_{ij}^1} \leq 1$ ,  $a = e^{c_{11}^1} - e^{c_{12}^1}$  and  $b = e^{c_{11}^2} - e^{c_{12}^2}$ , and this implies  $-1 \leq a \leq 1$  and  $-1 \leq b \leq 1$ , and as a consequence  $a-1 \leq 0$  and  $b-1 \leq 0$ . Then, if  $a \neq b$ , we have that  $z_1 = \frac{1-b}{a-b}$  and  $z_2 = \frac{a-1}{a-b}$  have opposite signs, and this not satisfy item (i).

Let us now analyze the case  $z_1 = \frac{p}{a}, z_2 = \frac{1-p}{b}$ . We will show that, either this solution does not satisfy item (i), or the dominant eigenvalue of the matrix in (17) is greater than 1, and therefore do not satisfy item (iii). In fact, the matrix in (17) is

$$B = \frac{p}{e^{c_{11}^1} - e^{c_{12}^1}} \begin{pmatrix} e^{c_{11}^1} & e^{c_{12}^1} \\ 1 - e^{c_{11}^1} & 1 - e^{c_{12}^1} \end{pmatrix} + \frac{1-p}{e^{c_{11}^2} - e^{c_{12}^2}} \begin{pmatrix} e^{c_{11}^2} & e^{c_{12}^2} \\ 1 - e^{c_{11}^2} & 1 - e^{c_{12}^2} \end{pmatrix}.$$

This matrix have eigenvalues given by  $\lambda_1 = 1$  and  $\lambda_2 = \frac{p}{e^{c_{11}^1} - e^{c_{12}^1}} + \frac{1-p}{e^{c_{11}^2} - e^{c_{12}^2}} = \frac{p}{a} + \frac{1-p}{b}$ . If this solution satisfy item (i), then  $0 < e^{-\varphi_1} = z_1 = \frac{p}{a}$  and  $0 < e^{-\varphi_2} = z_2 = \frac{1-p}{b}$ , and then we have  $0 < a \leq 1$  and  $0 < b \leq 1$ . This implies  $\frac{p}{a} + \frac{1-p}{b} \geq p + 1 - p = 1$ , (with equality only if  $a = b = 1$ , i.e.,  $z_1 = p$  and  $z_2 = 1-p$ ). In the case we have strict inequality,  $\lambda_2 > 1$  will be the dominant eigenvalue.

We conclude that  $(p, 1-p)$  is the only critical point that satisfy items (i), (ii), and (iii). Hence, as we know that there exist a solution to

$$\inf_{\varphi: P(c-\varphi)=0} \int_X \varphi(x) d\mu$$

it must be given by  $(\tilde{\varphi}_1, \tilde{\varphi}_2) = (-\log(p), -\log(1-p))$ , which finishes the proof of item b) of Theorem 6. ■

**Remark:** Note that the original problem is to minimize  $\int_X \varphi(x) d\mu = \varphi_1 p + \varphi_2 (1-p)$  subject to the restriction  $\{\varphi : P(c-\varphi) = 0\}$ . We know by [6], that exists a unique minimizer to this problem, that we denote by  $\tilde{\varphi} = (\tilde{\varphi}_1, \tilde{\varphi}_2)$ , and that satisfies  $P(c-\tilde{\varphi}) = 0$ . Hence we get that the matrix  $B_{\tilde{\varphi}} = e^{-\tilde{\varphi}_1} C^1 + e^{-\tilde{\varphi}_2} C^2$  (see equation (17)) has 1 as dominant eigenvalue and this eigenvalue is simple.

Instead of finding directly the minimizer of the original problem, we solved a second problem, which is to minimize  $f(z_1, z_2) = -p \log z_1 - (1-p) \log z_2$ , subject to  $g(z_1, z_2) = 0$ , where  $z_1 = e^{-\varphi_1}, z_2 = e^{-\varphi_2}$ .

**Claim:** The solution of the original problem  $\tilde{z} = (\tilde{z}_1, \tilde{z}_2) = (e^{-\tilde{\varphi}_1}, e^{-\tilde{\varphi}_2})$  can be found in the set of extremal points of the second problem, that we determine using Lagrange multipliers.

*Proof of the claim:*

The conic  $g(z) = 0$  is given by points whose associated matrices have 1 as one of its eigenvalues. Such conic can have two subsets: The first of them given by matrices having 1 as maximal eigenvalue (call  $C_1$  this subset) and other (call it  $C_2$ ) having 1 as minimal eigenvalue. ( $C_1$  and  $C_2$  can intersect in points whose associated matrices have both eigenvalues given by 1.) Let us show that  $\tilde{z}$  is the interior of  $C_1$ . In fact, first note that  $B_{\tilde{z}}$  is strictly positive, which implies the dominant eigenvalue of the matrix  $B_{\tilde{z}} = \tilde{z}_1 C^1 + \tilde{z}_2 C^2$  is simple, and therefore the other eigenvalue is smaller than 1. Now, using the fact that the spectrum of  $B_z = z_1 C^1 + z_2 C^2$  varies continuously in the parameter  $z$ , we know that there exists a neighbourhood  $V_{\tilde{z}}$  of  $\tilde{z}$  in the curve  $g(z) = 0$  such that the matrix  $B_z$  has 1 as dominant and simple eigenvalue for any  $z \in V_{\tilde{z}}$ . This implies that  $\tilde{z}$  is the interior of  $C_1$ . Finally, as  $f(\tilde{z}) < f(z)$  for all  $z \in V_{\tilde{z}}$  (because  $\tilde{z}$  is a minimum point of the original problem), this implies  $\tilde{z}$  is a local minimum of the second problem and can be found using Lagrange multipliers. *End of Proof of Claim.*

*Remark on the claim:* Note that the level curves of the function  $f(z_1, z_2)$  have a simple geometry:  $z_2$  is, in fact, proportional to a negative power of  $z_1$ , in any level curve. Also,  $g(z_1, z_2) = 0$  determines a conic, which also has a simple geometry, and this prevents any pathologies that could invalidate the argument above.

Now we want to perform Step 2.

The normalized cost associated to  $A = c - \tilde{\varphi}$  is

$$\bar{A} = A + \log(h_A) - \log(h_A \circ \sigma) - \log(\lambda_A),$$

where  $h_A$  is the eigenfunction of  $L_A = L_{c-\tilde{\varphi}}$  associated to the maximal eigenvalue  $\lambda_A = \lambda_{c-\tilde{\varphi}}$ . We see in the proof of the theorem 6 that  $\lambda_{c-\tilde{\varphi}} = 1$ . And using (14) we get that  $h_A = v_{c-\tilde{\varphi}}$  is the left eigenvector of  $B$  associated to the eigenvalue 1, where

$$B = \begin{pmatrix} e^{c_{11}^1} e^{-\tilde{\varphi}_1} + e^{c_{11}^2} e^{-\tilde{\varphi}_2} & e^{c_{12}^1} e^{-\tilde{\varphi}_1} + e^{c_{12}^2} e^{-\tilde{\varphi}_2} \\ e^{c_{21}^1} e^{-\tilde{\varphi}_1} + e^{c_{21}^2} e^{-\tilde{\varphi}_2} & e^{c_{22}^1} e^{-\tilde{\varphi}_1} + e^{c_{22}^2} e^{-\tilde{\varphi}_2} \end{pmatrix} = z_1 C^1 + z_2 C^2,$$

i.e.,  $h_A B = h_A$ , where  $(z_1, z_2) = (e^{-\tilde{\varphi}_1}, e^{-\tilde{\varphi}_2})$ .

**Remark:** If we are in case b) of Theorem 6, then the cost  $A = c - \tilde{\varphi}$  is already normalized. In fact, if we denote by  $p(1) = p = e^{-\tilde{\varphi}_1}$  and  $p(2) = 1 - p = e^{-\tilde{\varphi}_2}$ , using (19) we have

$$\sum_{x \in X} \sum_{\sigma(w)=y} e^{c(x,w)-\tilde{\varphi}(x)} = \sum_{x \in X} p(x) \sum_{a=1}^2 e^{c(x,a,y_1)} = \sum_{x \in X} p(x) [e^{c(x,1,y_1)} + 1 - e^{c(x,1,y_1)}] = 1,$$

for all  $y \in \Omega$ .

**Example:** Let us suppose that the cost  $c(x, y) = c(x, y_1, y_2)$  is represented by the following matrices

$$C^1 = \begin{pmatrix} 3 & 5 \\ 2 & 4 \end{pmatrix}, \quad C^2 = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}.$$

In order to perform Step 1 we look for positive solutions of (16), and we get that the unique positive solution is  $(z_1, z_2) = (0.101972, 0.0568922)$ .

Now in order to perform Step 2 we need first to calculate the left eigenvector, associated to the eigenvalue 1, of the matrix  $B = z_1 C^1 + z_2 C^2$ .

We have that

$$B = \begin{pmatrix} 0.4197 & 0.566751 \\ 0.431512 & 0.578563 \end{pmatrix}$$

and we can get that  $h_A = (0.596709, 0.802458)$  is such that  $h_A B = h_A$ .

Then the matrices

$$\bar{C}^1 = \begin{pmatrix} 0.3059 & 0.379132 \\ 0.274264 & 0.407887 \end{pmatrix}, \quad \bar{C}^2 = \begin{pmatrix} 0.113784 & 0.0423052 \\ 0.306036 & 0.170677 \end{pmatrix},$$

which are obtained by the expression

$$\bar{C}_{ij}^x = z_x C_{ij}^x \frac{h_A(i)}{h_A(j)} = e^{c(x,i,j)-\tilde{\varphi}(x)+\log(h_A(i))-\log(h_A(j))},$$

represent the normalized cost  $\bar{A} = A + \log(h_A) - \log(h_A \circ \sigma)$ .

In fact, by the equation (12), we have that  $C_{ij}^x = e^{c(x,i,j)}$ , as  $(z_1, z_2) = (e^{-\tilde{\varphi}(1)}, e^{-\tilde{\varphi}(2)})$  we have  $z_x C_{ij}^x = e^{c(x,i,j) - \tilde{\varphi}(x)}$ , and finally if  $y = (i, j, y_3, y_4, \dots)$  then  $\sigma(y) = (j, y_3, y_4, \dots)$ . Therefore

$$\bar{C}_{ij}^x = e^{c(x,i,j) - \tilde{\varphi}(x) + \log(h_A(y)) - \log(h_A \circ \sigma(y))} = e^{\bar{A}(x,y)}.$$

## References

- [1] J. Brémont. Gibbs measures at temperature zero. *Nonlinearity*, 16(2): 419–426, 2003.
- [2] J. H. Elton, An ergodic theorem for iterated maps. *Ergodic Theory Dynam. Systems*, v. 7, no. 4, pp 481-488, 1987.
- [3] B. Forte and F. Mendivil; A classical ergodic property for IFS: a simple proof. *Ergodic Theory and Dynamical System*, v. 18, no. 3, pp 609-611, 1998.
- [4] A. Katok and B. Hasselblatt; *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.
- [5] A. O. Lopes and J. Mengue; Duality Theorems in Ergodic Transport. *Journal of Statistical Physics*, v. 149, pp 921-942, 2012.
- [6] A. O. Lopes, J. Mengue, J. Mohr and R R. Souza; Pressure and Duality for Gibbs plans in Ergodic Transport. *Preprint 2014*, accepted for publication in the Bulletin of the Brazilian Mathematical Society.
- [7] K. Oliveira and M. Viana; *Fundamentos da Teoria Ergódica*, Sociedade Brasileira de Matemática, 2014.
- [8] W. Parry and M. Pollicott; Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque* Vol 187-188, 1990.
- [9] C. Villani; *Topics in optimal transportation*, AMS, Providence, 2003.

JOANA MOHR  
DEPARTAMENTO DE MATEMATICA  
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS  
AVENIDA BENTO GONÇALVES, 9500  
91509-900 - PORTO ALEGRE RS  
BRAZIL  
email joana.mohr@ufrgs.br

RAFAEL RIGÃO SOUZA  
DEPARTAMENTO DE MATEMATICA  
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS  
AVENIDA BENTO GONÇALVES, 9500  
91509-900 - PORTO ALEGRE RS  
BRAZIL  
email rafars@mat.ufrgs.br

The author RRS is partially supported by FAPERGS (proc.002063-2551/13-0).